Report LBNL-58024

Improvements to the image processing of HST NICMOS observations with multiple readouts

V. Fadeyev, G. Aldering and S. Perlmutter

Physics Division
Lawrence Berkeley National Laboratory
1 Cyclotron Rd, Berkeley, CA 94720;
VAFadeyev@lbl.gov, GAldering@lbl.gov, Saul@lbl.gov

ABSTRACT

We report on improvements made to the standard NICMOS processing pipeline. The calculation of the uncertainties on the signal accumulation rate has been modified to include the statistical correlations between the consecutive readouts. In order to correct a problem with the existing cosmic ray rejection algorithm, we have developed and implemented a joint fit procedure, where the accumulating signal is fit as linear functions of time with the same rate both before and after the cosmic ray (CR) impact. We also accounted for inter-pixel correlations in the CR-affected region. The new processing is most relevant for deep observations of faint targets, and for PSF fitting, for which unbiased measurements of accurate error estimates are important. We show examples of these improvements for deep NIC2 images of high-redshift supernova from the Supernova Cosmology Project.

Subject headings: methods: analytical — methods: data analysis — methods: statistical — space vehicles: instruments — techniques: image processing

1. Introduction

The Near Infrared Camera and Multi-Object Spectrometer (NICMOS) (Thompson *et al.* 1998) is one of the most successful instruments of the Hubble Space Telescope providing infrared images free of atmospheric influence. The instrument data have contributed to more than 100 publications in the last 5 years.

At the centerpiece of the instrument are HgCdTe infrared arrays manufactured by Rockwell Scientific. Imaging with the arrays is different from imaging with CCD-based devices in several aspects. The three features most relevant to this paper are: the existence of an operation mode with multiple non-destructive readouts (MULTIACCUM), relatively high readout noise ($30e^-$ versus $5e^-$ typical for the CCDs (Noll *et al.* 2004; Heyer *et al.* 2004)), and significant cosmic ray (CR) pollution caused by operation in space. The MULTIACCUM readout mode allows one to follow the time development of the signal in a given pixel. This information can be used in the linear fit to extract the source count rate, as is done in the NICMOS data processing pipeline [1]. The fit with multiple readouts effectively reduces the effect of the large readout noise, and gives better count rate estimate than the simple difference of the final and initial readouts (Garnett and Forrest 1993; Offenberg *et al.* 2001; Fixen *et al.* 2000; Thompson *et al.* 1998; Sparks 1998). The timing information also allows one to correct for the CR impact and potential-well saturation on a per-pixel basis.

In this paper we present improvements to the NICMOS pipeline processing, resulting in a better error estimate of the signal count rate and further suppression of the CR hits.

The rest of the paper is organized as follows. In Section 2 we point out the deficiency of the standard error determination and present an improved technique. In Section 3 we describe our method for a linear fit procedure for handling CR hits. In Section 4 we describe a way to avoid a bias in the pixels neighboring a CR-affected area. We show an example of the processing in Section 5, followed by our conclusions in Section 6.

---

[1] Throughout this paper we assume that the measured count rate is independent of the readout time. There are some recent indications to the contrary in the case of short exposures of faint objects (Bohlin 2005). The correction algorithm is not immediately obvious, due to the necessity to "linearize" the measured count rate on the per-quadrant basis due to deviations caused by spurious bias jumps. A future pipeline modification may correct for the effect after it is better characterized.

2. Linear fit and Poisson signal correlations

There are two components of the signal y_i read out from the detector during the ith readout: a Poisson count P_i due to photon sources, and the Gaussian readout noise r_i (Equation 1). The readout noise smears the source statistics. Throughout this paper we assume r_i to be un-correlated for different readouts, and to have constant standard deviation R during the exposure.

$$y_i = P_i + r_i \tag{1}$$

The signal from the photon sources results in correlation between the readouts, since the signal accumulated by the time of one readout affects the statistics of the following readouts. The correlation between readouts i and $i+k$ is

$$corr(i, i+k) = \frac{P_i}{\sqrt{P_i + R^2}\sqrt{P_{i+k} + R^2}} \tag{2}$$

The standard least-squares linear fit formulae for the count rate b are shown in Appendix A. They account for a weighting factor σ_i associated with readout y_i . It is natural to define the weight factor through the readout variance $\sigma_i = \sqrt{Var(y_i)} = \sqrt{P_i + R^2}$. It is this formalism that is implemented in the standard NICMOS pipeline.

As evident from Equation 2, the assumption of no correlations is not strictly true, and is violated to a degree dependent on the relative strength of the Poisson photonic source and the readout noise. In the limit where the photon counts are small compared to the readout noise, the correlation between the readouts vanishes, and the standard formulae shown in A1 become a reasonable approximation.

In the case of NICMOS there exists the phenomenon of "amplifier glow", where the amplifiers positioned at the corners of the four quadrants warm up during each readout becoming a source of thermal radiation to which the infrared detectors are sensitive. The effect results in deposition of 10-15 e^- signal in the center of the detector per readout, and an order of magnitude larger value in the corners. For long exposures with over 20 MULTIACCUM readouts the amplifier glow is more significant in the center of the detector than other sources of the background photons, such as dark current and zodiacal light, and its variance is comparable to that of the readout noise. In this case the correlations between the readouts become appreciable. The correlations are even more important for bright targets, and for objects imaged in the corners of the detector, due to the increased Poisson component of the readout signal.

Given the presence of correlations, we can attempt to improve upon the standard procedure. Monte Carlo simulations indicate, however, that including the correlations in the fit improves the accuracy of the count rate estimate by at most 15% for the sky-limited data in the corner of the arrays. The improvement is only 3% for the center of the array. This factor is marginal enough that we did not modify the formalism of the count rate derivation in the NICMOS pipeline. These simulations also indicate that the uncorrelated linear fit does not introduce a bias in the estimate.

However, the accuracy of the *error* on the count rate in the standard method is not good enough to be trusted. The count rate error derived according to Equation A1, $\sigma_{uncorr}(b)$, underestimates the true standard deviation, $\sigma_T(b)$. Figure 1 shows the ratio $\sigma_T(b)/\sigma_{uncorr}(b)$ as a function of the source rate. The dependence was obtained from Monte Carlo simulation with nominal input parameters: gain of 5.4 e^-/ADU, amplifier glow of 15 e^-/readout, readout noise of 27 e^-, dark current of 0.050 e^-/sec, MIF1024[2] readout sequence with 26 readouts. One can see that the error derived according to Equation A1 underestimates the true standard deviation by a factor of 1.4 for the sky pixels. For pixels with a source rate of 5 e^-/sec the factor is 2.9.

In the absence of the readout noise, the independent variables are the accumulated differences between the subsequent readouts. After rewriting the formulae via the differences $\delta y_i = P_i - P_{i-1} + r_i - r_{i-1} \simeq P_i - P_{i-1}$, one can estimate the part of the b variance which is due to the correlations. The part of the b variance due to the readout noise can be estimated separately, as an additional independent component. The formulae are shown in Appendix B. We note that this concept has been fully described by Sparks (1998), who derived formulae for the case of un-weighted data. The formulae were re-derived in Gordon *et al.* (2005) for Spitzer data analysis. We show them for completeness, and as a precursor to the more sophisticated case in the next section.

The Monte Carlo simulations verified the correctness of the formulae B1-B5. To check the performance of the error estimates on the real data, we histogramed the value of $b/\sigma(b)$. We examined the images with flat sky background and a small number of source objects. After the sky subtraction, the histogram for such images should be close to a Gaussian with unit width, if the derived errors reflect the true scatter of the sky fluctuations. For the data taken with MIF1024 and SPARS64 readout sequences we see that the distribution is close to Gaussian, with a width too narrow by the factor of 1.14 (Figure 2). We consider this to be a big improvement compared to the factor of 1.42 obtained with the same data using the old formulae (A1). The histograms for both cases are shown in Figure 2. The 14% deviation from unity could be due to a number of reasons, including the accuracy of calibration, the assumptions we made about the properties of the readout noise, and the count rate ramp-up effect discussed in Bohlin (2005).

3. Pixels Affected by Cosmic Rays

As mentioned in the introduction, the existence of multiple non-destructive readouts allows one to better account for cosmic ray hits. In the case of a cosmic ray hit, there is a jump in the signal accumulation in an affected pixel. It can be identified as a jump in the consecutive readout differences normalized to the time between the readouts. In the NICMOS pipeline, the identification

[2](Noll *et al.* 2004)

is performed via the differences between the readout values before and after the candidate jump and the linear fit: $(y_{i+1} - (a + b \times t_{i+1}))/\sigma_{i+1} - (y_i - (a + b \times t_i))/\sigma_i$. The default threshold for identifying CRs is 4σ.

The standard procedure in case of a CR hit is to shift the data following the jump on the basis of the two readout values straddling the CR hit, $\delta(y_{i+1}) = y_{i+1} - y_i$, and then refit the new sequence of data to the linear function using formulae A1.

Some of the pixels affected by cosmic rays and processed according to this procedure can still be visually identified in images as outliers. We attribute this feature to the finite precision of the jump measurement. The readout noise contribution can make $\delta(y_{i+1})$ differ from the "true" value of the CR deposition by an amount comparable to the standard deviation of the readout noise. The difference systematically shifts the values of all post-CR readouts from what would have been an un-biased estimate in the absence of the CR. This affects both the count rate determination and its error.

To avoid the CR processing effect described above, we developed a joint fit procedure, whereby both the readouts before and after the CR jump are fit to linear functions with the same slope: $y_1 = a_1 + b \times \mathit{time}$; $y_2 = a_2 + b \times \mathit{time}$. In this way the fit naturally accounts for the jump $\delta(y) = a_2 - a_1$ on the basis of all available readouts, and there are no artificial shifts in the data. The exact formulae are presented in Appendix C.

We note that there is also an alternative method used in Gordon *et al.* (2005), where the count rate is estimated separately from each of the readout intervals of the readout sequence (partitioned by the CRs), and then the measurements are combined. Our method gives a similar result, but it might be slightly more precise due to the postulate of the same count rate value for different intervals.

One could attempt to modify the CR identification procedure by using the joint fit at a hypothetical CR impact time, and using $\delta(y) = a_2 - a_1$ divided by its error as a measure of the jump. Such an algorithm was included in the Spitzer instruments pipeline described in Gordon *et al.* (2005). Our attempt to implement such a procedure indicates a possible bias, in which the effect of a spurious fluctuation is exaggerated, and an artificial positive slope is introduced in the fit. The effect is visually noticeable, perhaps because of the time sampling involved. For a MIF1024 sequence, there are only 8 readouts in the middle of the sequence, where a CR is most likely to occur (there are other "fast" readouts at the very beginning and very end of the sequence). For this reason, we have not modified the original CR identification algorithm.

We also note that CRs sizably increase the count rate error for sky pixels. For the worst-case scenario with a CR occurring in the middle of the exposure, the error increases by about a factor of two, due to the decreased time axis range in the fit (two halves combined are worse than one whole). The "shifting" procedure in the standard NICMOS pipeline does not account for this effect.

4. CR Neighborhood Pixels

On examining the processed images we noticed one additional artifact: a number of “sky” pixels adjacent to the CR-affected regions visually appear to be positive outliers. The readout sequence of the outlier pixels show moderate (below-threshold) jumps at the same time as the nearby CR-affected pixels. The effect is illustrated in Figure 3 showing the correlation between the same-time jumps for the neighboring pixels. We note that the correlation is obvious for the side neighbors, but not for the corner neighbors. We show an example of such influence in the real exposure in the Figures 4 and 5.

We attribute this phenomenon to the CR particles interacting with the array material, and possibly spawning secondary particles, such as delta electrons and bremsstrahlung radiation.

To remove the bias, we process the images in two passes. The CR-identification algorithm is run during the first pass. The time locations of the CR jumps are flagged for each pixel in the array. Then the flags are propagated to the same time locations for the side neighbors. Finally, the second pass of the algorithm is run to refit the data while taking into account the previously identified jumps, and the count rates are extracted for all pixels. This procedure drastically reduces the number of outliers remaining in the images.

5. Processing example

We show an example of the effect of our processing in Figures 6a and 6b. By visual examination of the same exposure processed with the standard pipeline and the modified version, one might conclude that the number of positive outliers is reduced, but that some negative outliers were also introduced. However, we caution against conclusions based on the pixel count rates only; the uncertainty on the rate is an equally important scale ingredient in deciding whether a pixel is an outlier. As we mentioned in Section 3, the error necessarily increases for the pixels affected by CRs. (This is one of the arguments for using the measured count rate error information in all photometric procedures.)

It is instructive therefore to take a look at the images where the sky-subtracted count rates are divided by their estimated uncertainties. As shown in Figures 6c and 6d, there is a clear improvement in the case of the new (modified) NICMOS pipeline. The new image is very uniform, indicating that the pixels which look like outliers in the count rate images have correctly estimated uncertainties, and are therefore consistent with the sky level on that scale.

6. Conclusions

We have improved the NICMOS pipeline processing in three areas: 1) We made the count error estimates more reliable, 2) We improved the CR rejection procedure, 3) We have accounted for biases in the pixels neighboring cosmic rays.

The improvements are most relevant for analyses which include the count rate uncertainties, and for observations of faint objects.

The relevance of our improvements for future space-based infrared instruments, such as those for JWST or for JDEM, depends on the amount of the readout noise the infrared arrays possess. Larger noise calls for more consideration to be given to the pipeline processing, and for more readouts during an exposure.

We are grateful to Howard Bushouse, Chris Bebek, Al Schultz, and Ralph Bohlin for discussions regarding the NICMOS features, performance, and calibration.

Partial support for this work was provided by NASA through grants from the Space Telescope Science Institute (STScI), associated with program numbers 8088, 9075. STScI is operated by Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555. This work was also supported in part by the Director, Office of Science, Office of High Energy and Nuclear Physics, of the U.S. Department of Energy under Contract No. DE-AC03-76SF000098.

A. Classic Linear Fit Formulae

Equations A1-A4 are the linear fit formulae for uncorrelated data. The dependence of signal (photon counts) y_i on the readout time x_i is fit to the linear dependence y_i and the count rate b is derived.

$$b = \frac{1}{Det}(S \times S_{xy} - S_x \times S_y) \tag{A1}$$

$$\sigma(b) = \frac{1}{\sqrt{Det}} \tag{A2}$$

$$Det = SS_{xx} - S_x^2 \tag{A3}$$

$$S = \sum_{i=1}^{n} \frac{1}{\sigma_i^2}, \quad S_x = \sum_{i=1}^{n} \frac{x_i}{\sigma_i^2}, \quad S_{xx} = \sum_{i=1}^{n} \frac{x_i^2}{\sigma_i^2}, \quad S_y = \sum_{i=1}^{n} \frac{y_i}{\sigma_i^2}, \quad S_{xy} = \sum_{i=1}^{n} \frac{x_i y_i}{\sigma_i^2} \tag{A4}$$

B. Count Rate Error for the Case of Inter-Readout Correlations

Equations B1-B5 present the count rate estimate accounting for the inter-readout correlations.

$$\sigma(b) = \frac{1}{\Delta}\sqrt{\sigma(P)^2 + \sigma(G)^2} \tag{B1}$$

$$\Delta = S_{xx} - S_x^2/S \tag{B2}$$

$$\sigma(P)^2 = \sum_{k=2}^{n} S_k^2 \left(\frac{S_x}{S} - \frac{S_{xk}}{S_k}\right)^2 \frac{\delta y_k}{gain} \tag{B3}$$

$$\sigma(G)^2 = R^2 \sum_{k=1}^{n} \frac{1}{\sigma_k^4} \left(\frac{S_x}{S} - x_k\right)^2 \tag{B4}$$

$$S_k = \sum_{i=1}^{k-1} \frac{1}{\sigma_i^2}, \quad S_{xk} = \sum_{i=k-1}^{n} \frac{x}{\sigma_i^2} \tag{B5}$$

Here the $\sigma(P)^2$ term comes from the Poisson part of the readout values, and $\sigma(G)^2$ is due to the Gaussian readout noise.

C. Count Rate Error for Pixels Affected by the Cosmic Rays

Here we fit both the readout sequences before and after a CR jump to linear functions with the same slope: $y_1 = a_1 + b \times time; y_2 = a_2 + b \times time.$

For the case of a single CR jump we can define the joint χ^2 as following.

$$\chi^2 = \sum_{i=1}^{k} \frac{(y_i - (a_1 + b \times x_i))^2}{\sigma_i^2} + \sum_{i=k+1}^{n} \frac{(y_i - (a_2 + b \times x_i))^2}{\sigma_i^2} \tag{C1}$$

In case of N CR jumps the χ^2 will have $N + 1$ similar terms.

After taking derivatives of a_1 , a_2 , and b to minimize the χ^2, and solving the system of linear equation, we arrive at the following formulae for the count rate:

$$b = \frac{1}{\Delta} \sum_i (S^i \times S_{xy}^i - S_x^i \times S_y^i) \tag{C2}$$

$$\Delta = \sum_i (S_{xx}^i - (S_x^i)^2/S^i) \tag{C3}$$

The summation in Equations C2, C3 runs over the intervals separated by the CR jumps. For the case of a single interval (no CR jump), the expression for b reduces to A1. For $\sigma(b)$ we derive an expression similar to B1:

$$\sigma(b) = \frac{1}{\Delta}\sqrt{\sum_i (\sigma(P)_i^2 + \sigma(G)_i^2)} \tag{C4}$$

$$\sigma(P)_i^2 = \sum_{k=2}^{n_i} (S_k^i)^2 \ \left(\frac{S_x^i}{S^i} - \frac{S_{xk}^i}{S_k^i}\right)^2 \ \frac{\delta y_k}{gain} \tag{C5}$$

$$\sigma(G)_i^2 = R^2 \sum_{k=1}^{n_i} \frac{1}{\sigma_k^4} \ \left(\frac{S_x^i}{S^i} - x_k\right)^2 \tag{C6}$$

REFERENCES

Bohlin, R., Lindler, D., and Riess, A., Instrument Science Report NICMOS 2005-002, (Baltimore:STScI).

Garnett, J.D. and Forrest, W.J., 1993, Proc. SPIE, 1946, 395.

Gordon, K.D. *et al.*, 2005, PASP, 117, 503G.

Fixen, D.J., *et al.*, 2000, PASP, 113, 1350.

Heyer, Biretta, *et al.*, 2004, WFPC2 Instrument Handbook, Version 9.0, (Baltimore:STScI).

Noll, K., *et al.*, 2004, NICMOS Instrument Handbook, Version 7.0, (Baltimore:STScI).

Offenberg, J.D., *et al.*, 2001, PASP, 113, 240.

Sparks, W.B., Instrument Science Report NICMOS 98-008, (Baltimore:STScI).

Thompson, R.I., Rieke, M., Schneider, G., Hines, D.C., Corbin, M.R., 1998, ApJL, 492, L95.

This preprint was prepared with the AAS LaTeX macros v5.2.

Fig. 1.— Ratio of the true standard deviation of the count rate and the one derived according to the linear fit formulae A1, as a function of the signal rate.

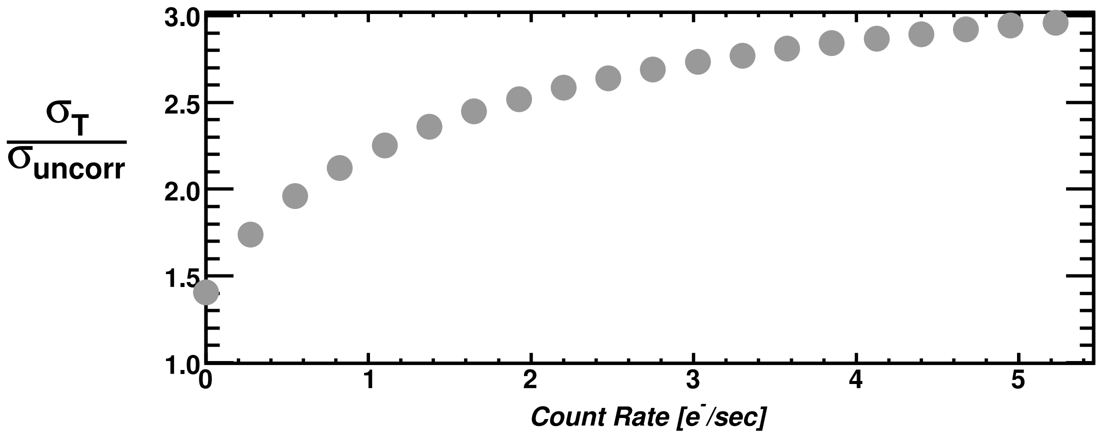
Fig. 2.— The distribution of the pixel count rate divided by its estimated uncertainty in a sky-dominated image after sky subtraction. The solid lines correspond to new error estimates accounting for the correlations between different readouts. The dashed line corresponds to the default error estimates in the NICMOS pipeline. The histogram on the right is the same as the one on the left, except that it is plotted on the logarithmic scale.

Fig. 3.— The sequential readout difference (in units of the standard deviation) in the neighboring pixel versus the primary pixel. Top: side neighbors. Bottom: corner neighbors. For visibility purposes, the left half of the X axis is shown in white color in these plots.

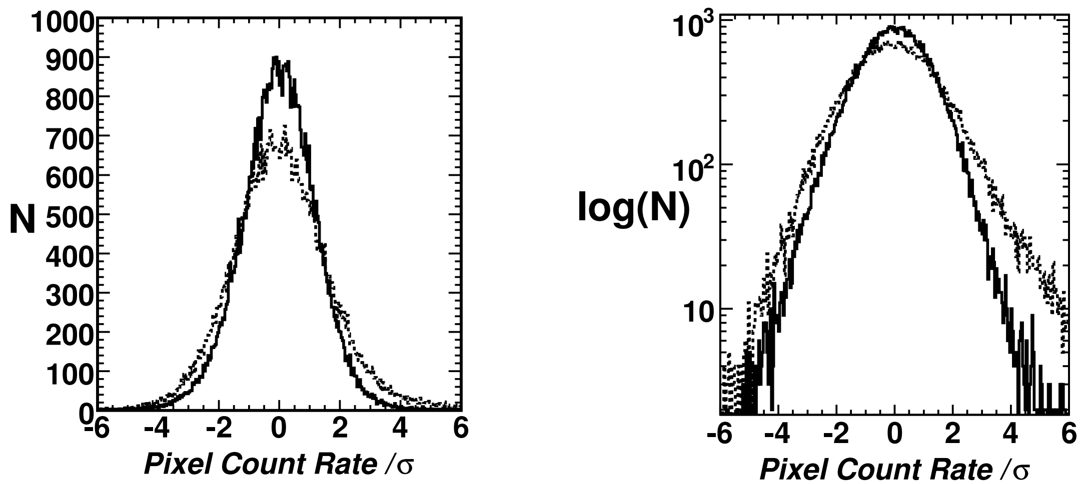
Fig. 4.— A part of reprocessed last readout frame from a NICMOS NIC2 exposure. The darker pixels indicate a larger counts, which could be due to cosmic rays. The dark region at the bottom is a field galaxy. The selected 3x3 pixel box is used as an example in the next Figure.

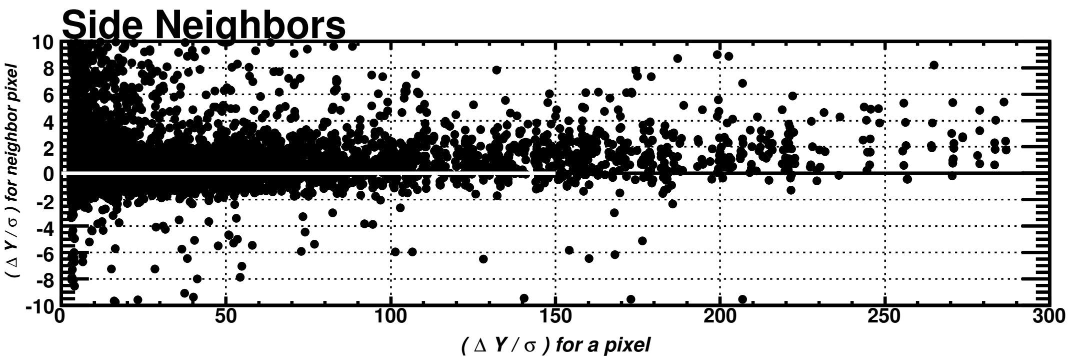
Fig. 5.— The time development of the counts in the 3x3 box selected in Figure 4. The numbers in the figure show the calculated significance of the jump at ≈350 sec. The default threshold is 4σ. Note that only one of the side neighbors of the central pixel is above this threshold (labeled "South Neighbor" on the plot subpanel).

Fig. 6.— NICMOS NIC2 count rate image of a supernova target. The darker pixels correspond to higher count rates. Image 6a is obtained with the default NICMOS pipeline processing, and the image 6b is obtained with our improved processing. Images 6c and 6d correspond to 6a and 6b, except that the sky-subtracted count rates were divided by their estimated uncertainties. The supernova is near the center of the field. There are two faint field galaxies, one is on the right hand side of the picture, and another is at about the same distance directly below the supernova. We vetoed pixels along 128th row, 128th and 129th columns, and at coronagraphic hole location from processing. This created visual peculiarities in the images: middle horizontal and vertical lines, and a circle circle in the upper left quadrant. The 45° line in the lower left quadrant is a diffraction spike from a star outside the field.

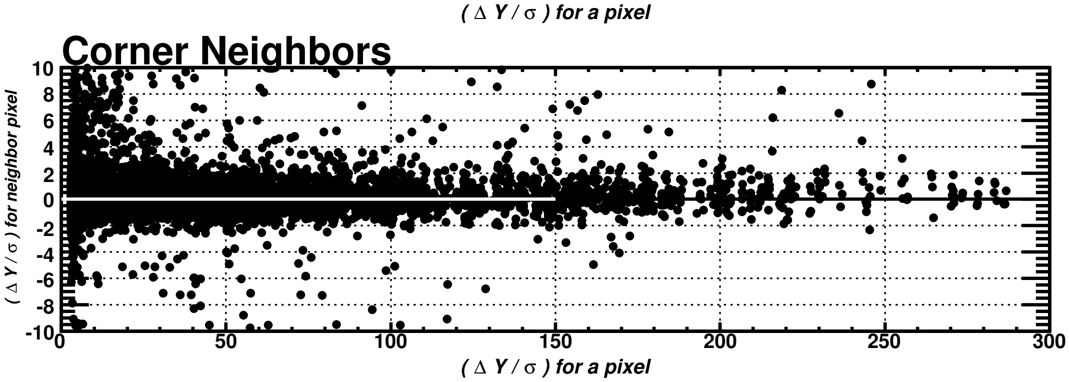


$\frac{\sigma_T}{\sigma_{uncorr}}$
3.0
2.5
2.0
1.5
1.0
0
1
2
3
4
5
Count Rate [e⁻/sec]

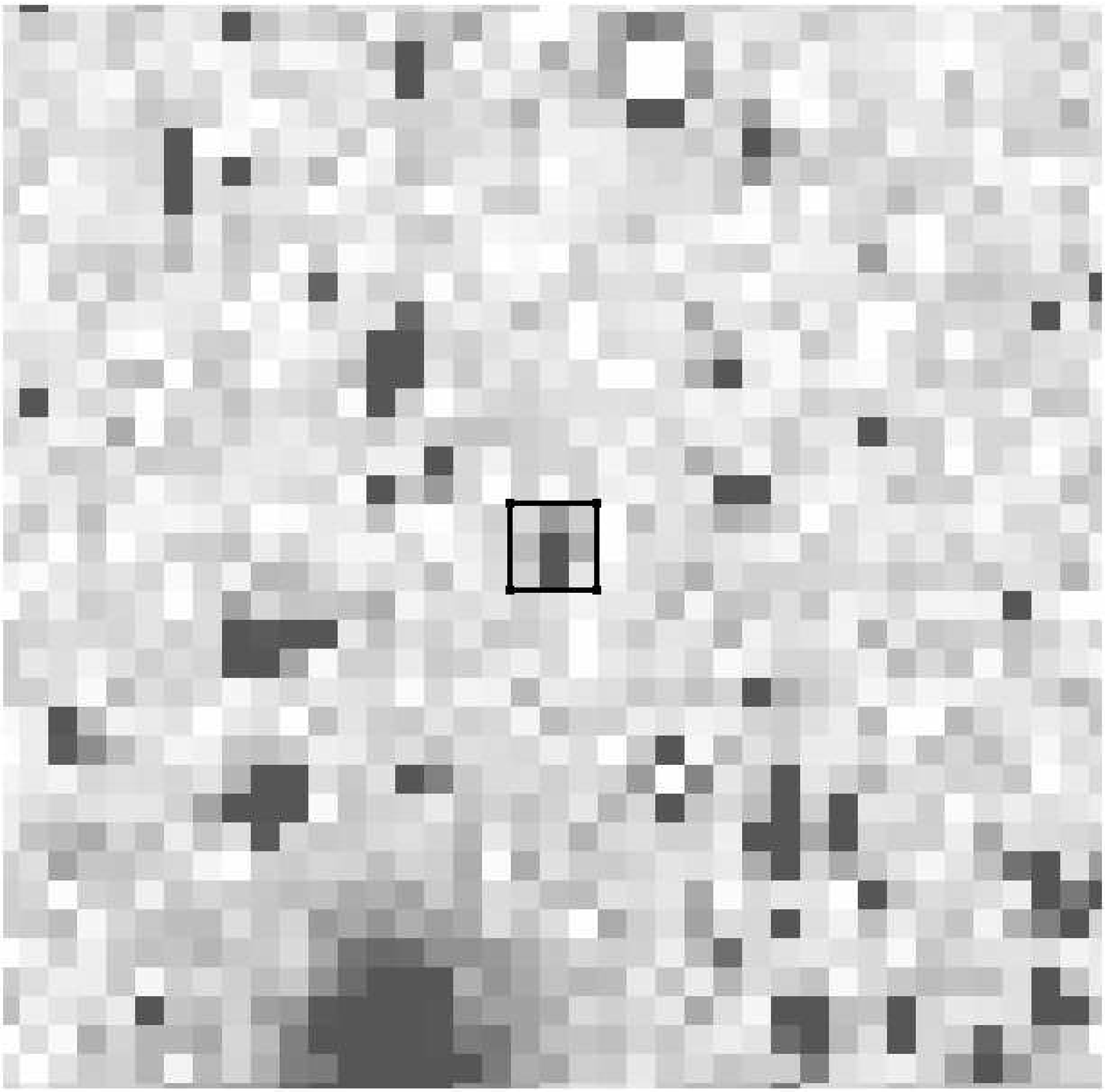

1000
900
800
700
600
500
400
300
200
100
0
N
-6
-4
-2
0
2
4
6
Pixel Count Rate /σ
10^3
10^2
10
log(N)
-6
-4
-2
0
2
4
6
Pixel Count Rate /σ

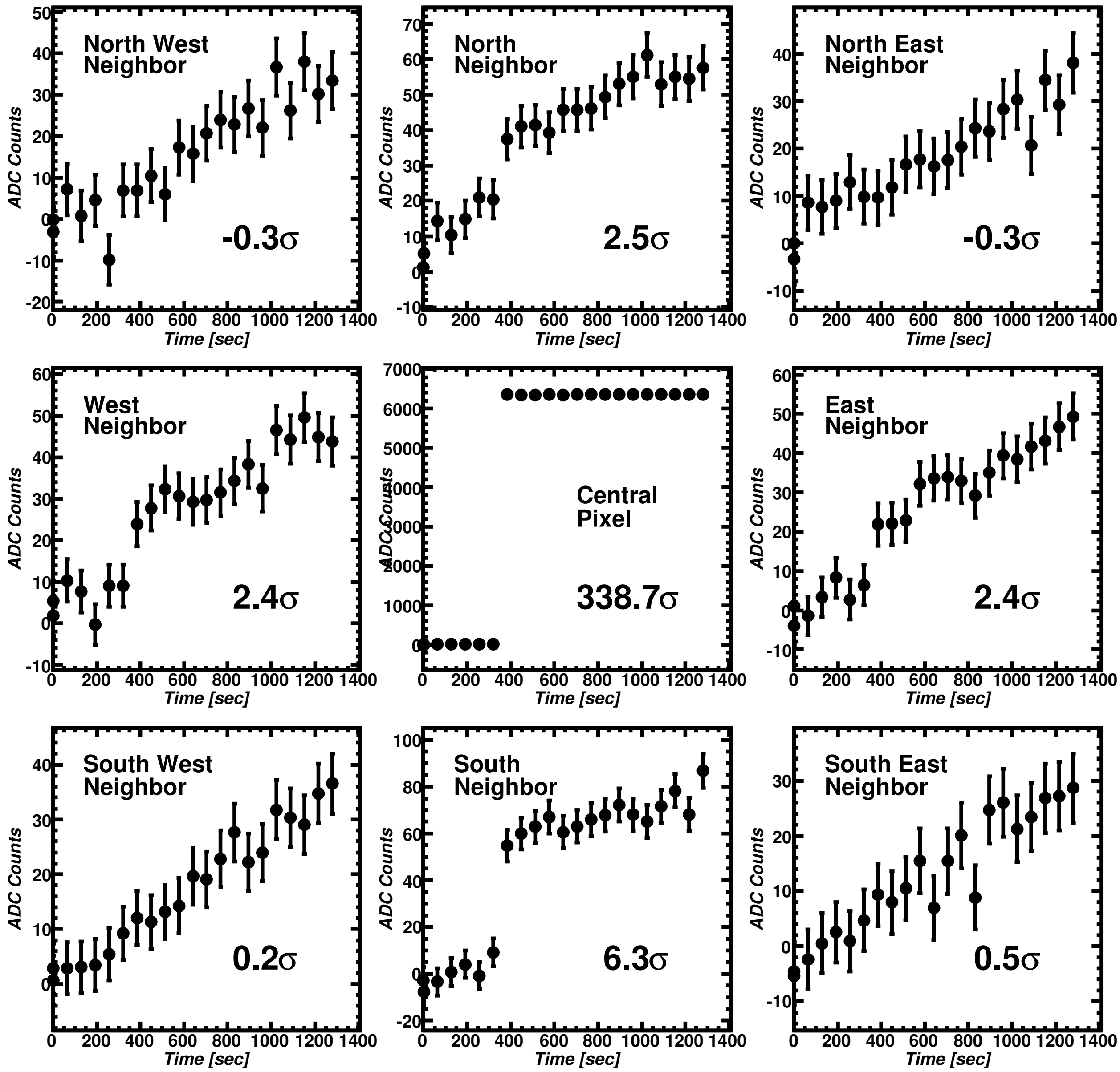


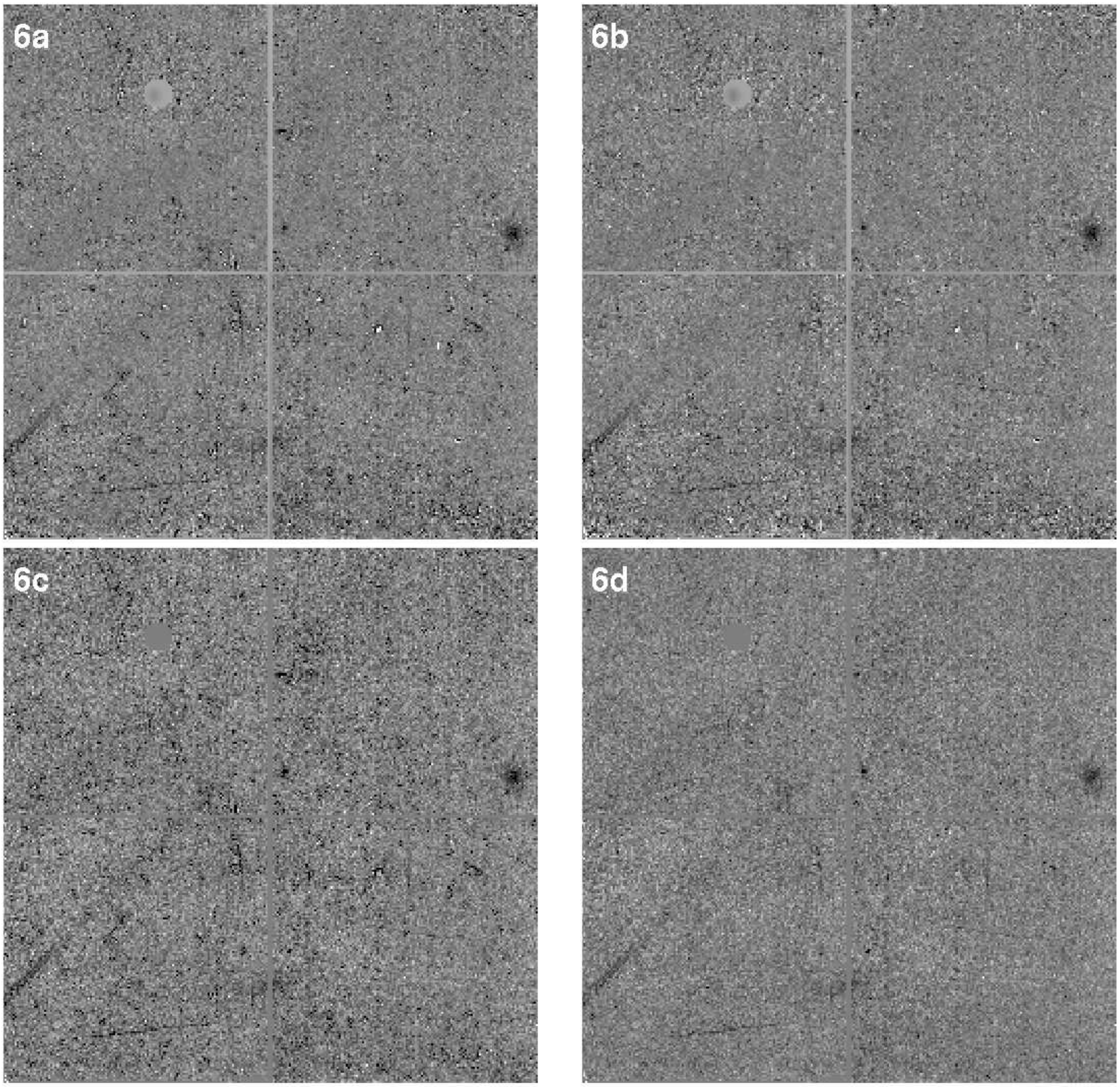
Side Neighbors
(Δ Y / σ) for neighbor pixel
(Δ Y / σ) for a pixel



Corner Neighbors
(Δ Y / σ) for neighbor pixel
(Δ Y / σ) for a pixel




North West Neighbor
-0.3σ
North Neighbor
2.5σ
North East Neighbor
-0.3σ
West Neighbor
2.4σ
Central Pixel
338.7σ
East Neighbor
2.4σ
South West Neighbor
0.2σ
South Neighbor
6.3σ
South East Neighbor
0.5σ
ADC Counts
Time [sec]


6a
6b
6c
6d